3. Graphical Overview of Optimization Theory (cont)

3.4. Constrained Maximization: One Variable.

**Question:** So far, we’ve been declaring that a necessary condition for a local maximum at $x^*$ is that $f'(x^*) = 0$. That is, if the slope isn’t zero at $x^*$ then I know I don’t have a local maximum at $x^*$. Now that was true, because of the way in which I defined the function $f$, but it was only true given this caveat. What was the critical part of the definition of $f$? In other words, under what conditions can a differentiable function $f$ have a local maximum at $x^*$ but the slope isn’t zero at $x^*$?

**Answer:** Whenever you are looking for a maximum on, say, an interval of the real line, the above conditions needn’t hold. In this case, we say that we are looking for a *constrained* maximum. The constraint is the fact that the maximum you are looking for must lie in a specific set.

Typically, in economic problems, the maximization problems we have to solve are constrained maximization problems. Very rarely, are economic problems completely unconstrained.

1. Most obvious constraint is that in many cases, the variable that you are maximizing must be non-negative (i.e., positive or zero).
(2) Might also be the case that you have bounds on either end of the problem, e.g., find the best location for a gas station between San Francisco and Los Angeles.

Consider Fig. 3.

(1) The graph on the left is defined on the interval \([x, \bar{x}]\). It has three local maxima but only for one of them is the slope zero. What are the conditions?

   (a) A necessary condition for \(f\) to attain a maximum at \(\bar{x}\) is that the slope of \(f\) at \(\bar{x}\) is zero or negative.

   (b) A necessary condition for \(f\) to attain a maximum at \(\bar{x}\) is that the slope of \(f\) at \(\bar{x}\) is zero or positive.

   (c) A necessary condition for \(f\) to attain a maximum at \(x^*\) strictly between these points is that the slope of \(f\) is zero.

(2) The graph on the right is defined on the interval \([0, \infty)\). In this case, we only have to worry about the point zero and all other points. What are the conditions?

   (a) A necessary condition for \(f\) to attain a maximum at 0 is that the slope of \(f\) at zero is nonpositive.

   (b) A necessary condition for \(f\) to attain a maximum at a positive number \(x^*\) is that the slope of \(f\) at \(x^*\) is zero.

These conditions are known as the Kuhn-Tucker necessary conditions for a constrained maximum. (Actually, they are known as the Karush-Kuhn-Tucker conditions. Karush was a graduate student who was involved in developing the conditions, but his name has been lost in history.)
3.5. **Unconstrained Maximization: Several Variables.**

Fig. 4 above is the graph of a function $f$ which has a nice global maximum in the interior of the domain. We are interested again in the relationship between the “slope” of the function $f$ at a point $x$ and the extrema (maxima and minima) of the function.

In this case, it’s less clear what we mean by “the slope” of a function of several variables.

- If the function has $n$ variables, then the slope of the function is an $n$-component vector.
- Each component of the vector denotes the “slope” of the function in a different “direction.”
• That is, slice the graph through the maximum point, (a) in a direction parallel to the horizontal axis, and (b) in a direction parallel to the vertical axis.
  – Each time you do this, you get a single variable function.
  – The slope you get from (a) is the first component of the vector of slopes.
  – It’s also known as the first partial derivative of the function with respect to the first variable.
  – The slope you get from (b) is the second component of the vector of slopes.
  – It’s also known as the first partial derivative of the function with respect to the second variable.

• The vector itself is sometimes called the derivative of \( f \) at \( x \). Also it is referred to as \textit{the gradient of} \( f \), evaluated at \( x \). It is denoted by \( f'(x) \), where the bold symbols denote vectors. In other words, the notation used for the slope, the first derivative and the gradient of \( f \) is the same.

Now we can go back to the issue of unconstrained maximization.

\textbf{Question:} Suppose I know that \( f \) attains an (unconstrained) local maximum at \( x^* \), i.e., a maximum at an interior point of the domain of \( f \). What can I say then about the slope of \( f \) at \( x^* \), i.e., the vector \( f'(x^*) \)?

\textbf{Answer:} The slope of \( f \) has to be zero at \( x^* \), i.e., \( f'(x^*) = 0 \). In this case, the statement \( f'(x^*) = 0 \) says that the whole vector is zero.

\textbf{Note:} There are lots of directions and the derivative only picks two of them, i.e., the directions parallel to the axes. It turns out that if certain conditions are satisfied— a sufficient (but not necessary) condition is that each of the partials is \textit{continuous function}—then knowing that the two partial derivatives are zero guarantees that the slope of the function is zero in \textit{every} direction. To see that this isn’t true generally, consider the function

\[
 f(x,y) = \begin{cases} 
 0 & \text{if } |x| = |y| \\
 \frac{xy}{x^2 - y^2} & \text{otherwise}
\end{cases}
\]

graphed in Fig. 6. The partial derivatives of this function are well defined for all \( x,y \), provided \( |x| \neq |y| \), and when either \( x \) or \( y \) is zero, both derivatives are zero. However, clearly you don’t have a max/min or anything else at zero. In the example \( f \) graphed in Fig. 6, the fact that the partials are both zero doesn’t guarantee that “derivatives in other directions” will be zero.

The problem with this example is that in a neighborhood of zero, the derivatives of \( f \) aren’t differentiable functions. To demonstrate this, Fig. 7 plots \( f_x(\cdot) \) as we go around a circle centered at zero, with very small

\textsuperscript{1} Marsden and Tromba, Vector Calculus, Third Edition, Theorem 9, p.172
Figure 6. Zero partials don’t necessarily guarantee zero slopes in other directions.

Figure 7. Plot of $f_x(\cdot)$ going around a circle of radius $\delta$.

Radius $\delta > 0$. On the horizontal axis we measure the angle $\theta$ between $x$ and the vector $(1, 0)$. The diagonals occur at $45^\circ$, $135^\circ$, etc. As the figure illustrates, the derivative is well defined everywhere except along the diagonals, but since the same picture holds for all positive $\delta$, the function is not continuously differentiable in a neighborhood of zero.

Summary: A necessary condition for $f$ to attain a local maximum at $x^*$ in the interior of its domain is that the slope of every cross-section of the function is zero. If each of the partial derivatives of $f(\cdot)$ is continuous at $x^*$, then zero slopes in the directions of each axis implies zero slopes in every direction.
Knowing that the slope of every cross-section is zero at $x^*$ does not in general imply that $f(\cdot)$ attains a local maximum at $x^*$. Fig. 8 below is an example in which the graph of the function is flat at $x^*$, i.e., the slopes of all of the cross-sections are zero, but you don’t have a local maximum at $x^*$. In this example, you have what’s called a saddlepoint of the function.

**Question:** Suppose I know that the slope of $f$ is zero at $x^*$. What additional information about the slope of $f$ at $x^*$ do I need to know that $f$ attains a strict local maximum at $x^*$?

**First shot at an answer:** A first guess is that the slopes of all of the partial derivatives have to be decreasing, i.e., the second order partials must all be negative. This condition would rule out the possibility of a saddle point. Turns out however that it isn’t enough. Reason is that just looking at partials doesn’t give you information about what’s going on with the other directional derivatives. Imagine a graph where if you took circular cross-sections, i.e., sliced the graph with a circular cookie-cutter, what you got when you laid out the cut was a sine curve, with zeros corresponding to the points at the axes. Here you might think you had a weak local maximum if you just looked at the second-order partials, but you can’t put the board on top.

*Figure 8. Saddle Point of a function of two variables*
More concretely, consider Fig. 9, which is the graph of the function \( f(x,y) = \begin{cases} 0 & \text{if } x = y = 0 \\ \frac{(xy)^2}{x^2+y^2} - 0.1(x^2+y^2) & \text{otherwise} \end{cases} \). To see what’s going on with this function, we’ll do an exercise similar to the one we did for Fig. 6, except that instead of plotting partial derivatives, we’ll look at the second-derivatives of the diagonal cross-sections as we work our way around the unit circle. More precisely, for each direction \( h \in [0,360^\circ) \), we look at \( f_{hh}(h_1,h_2) \), which is the directional derivative in direction \( h \) of the directional derivative in direction \( h \) of \( f \). (This hideous but unavoidable terminology will become familiar later on.) Observe that along the directions parallel to the axes, i.e., when \( \theta \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ, 360^\circ\} \), the slope of the slope of the diagonal cross-section is negative, while along directions parallel to the 45 degree lines, i.e., when \( \theta \in \{45^\circ, 135^\circ, 225^\circ, 315^\circ\} \), it is positive. For a negative definite function, this graph would be below the origin for all values of \( \theta \). Note well that this picture is not a pathological one like Fig. 6. All of the derivatives in sight are nicely continuously differentiable. So, we’ve established that negativity of the second partials, i.e., \( f_{ii} < 0 \), for each \( i = 1, \ldots, n \), isn’t enough to guarantee a maximum. So what do we need? The requirement is that the second derivative of the function \( f \) evaluated at \( x^* \) (which is a matrix) is negative definite. What does this mean? Mathematically a bit complex, but diagrammatically it means that you can put a flat board on top of the graph of the function at \( x^* \) and the graph will be everywhere below the board. Note that in figure 9, you can’t do this. Similarly, to establish that \( f \) attains a local minimum at \( x^* \), need to show that you can put a flat board below the graph of the function at \( x^* \) and the graph is everywhere above the board.
**Terminology:** Confusion always arises concerning the term second partial. Could mean the second of the first partial derivatives, i.e., the partial derivative w.r.t. to the second element of the domain. I’ll try to be consistent about the other usage: i.e., a second partial to me will (almost) always refer to a partial derivative of a partial derivative.

**Summary:** If first partial derivatives are all zero, then unless the function is pathological/nasty/not continuously differentiable, derivatives in all directions will also be zero. But if all the second partial derivatives have the same sign, then there’s nothing pathological about a function in which the second partial in some direction not parallel to the axis has a different sign.

In the summary table below, \( i \)'s denote components of \( \mathbb{R}^n \), i.e., \( i = 1, \ldots, n \), while \( h \) denote directions (infinitely many).

<table>
<thead>
<tr>
<th>Conditions for a Strict Local Maximum at ( x^* ).</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f : \mathbb{R} \to \mathbb{R} )</td>
</tr>
<tr>
<td>You need</td>
</tr>
<tr>
<td>First Order Condition</td>
</tr>
<tr>
<td>Second Order Condition</td>
</tr>
</tbody>
</table>

Local vs Global Optima: Once again, the first and second derivatives of the function at \( x^* \) tell you nothing in general about whether or not a local extremum of the function at \( x^* \) is a global extremum. There are conditions that we can impose on the function that does guarantee that a local extremum is either a global maximum or a global minimum.

**Question:** When the function \( f \) satisfies a certain property then knowing \( f'(x^*) = 0 \) is **SUFFICIENT** to conclude that \( f \) attains a **GLOBAL** maximum at \( x^* \). What is that property?

**Answer:** Same as before. A function is concave if the set of points that lie below the graph of the function is a convex set. Alternatively, the matrix of second partial derivatives (the Hessian) has to be everywhere negative semi definite. Once again, to see the role of concavity, consider the surface of a camel. Each hump is a local maximum, but one of them isn’t a global one unless the camel has perfectly balanced humps. Can’t
put a flat board on top of both humps of the camel and have the camel be underneath the board. Same thing for minima.

**Note well:** Earlier, I made a big deal of the fact that $f''(x^*) \leq 0$ was not a sufficient condition (along with $f'(x^*) = 0$) for a max, and gave the counterexample $f(x) = x^4$, which has the property that $f'(0) = f''(0) = 0$ but you have a strict global minimum at 0. Yet, here, I don’t require strictness, only weak concavity. The reason for the difference is that in the paragraph above, I’m stipulating (at least for a function with one argument) that $f''(\cdot) \leq 0$, i.e., that the second derivative is *everywhere* nonpositive, which is a much stronger condition.

**Fact:** If $f$ is a differentiable concave function, then a *necessary and sufficient* condition for $f$ to attain a *global* maximum at $x^*$ in the interior of its domain is that $f'(x^*) = 0$. Note that this is a *vector* equality.

**Fact:** If $f$ is a convex function, then a *necessary and sufficient* condition for $f$ to attain a *global* minimum at $x^*$ in the interior of its domain is that $f'(x^*) = 0$. Once again, this is a *vector* equality.

### 3.6. Level Sets, upper and lower contour sets and Gradient vectors

Next step is constrained maxima with several variables, but first need some terminology.

You all presumably know what contour lines are on a map: lines joining up the points on the map that are at the same height above sea level. Same thing in math, except slightly different terms. Three terms that you need to know:

1. **Level set:** A level set of a function $f$ consists of all of the points *in the domain of $f$* at which the function takes a certain value. In other words, take any two points that belong to the same level set of a function $f$: this means that $f$ assigns the same value to both points.

2. **Upper contour set:** An upper contour set of a function $f$ consists of all of the points *in the domain of $f$* at which the value of the function is *at least* a certain value. We talk about “the upper contour set of a function $f$ corresponding to $\alpha$”, referring to the set of points to which $f$ assigns the value at least $\alpha$.

3. **Lower contour set:** A lower contour set of a function $f$ consists of all of the points *in the domain of $f$* at which the value of the function is *no more than* a certain value.
For example, consider Fig. 10 below. The level sets are indicated on the diagram by dotted lines. Very important fact that everybody gets wrong: the level sets are the lines on the horizontal plane at the bottom of the graph, NOT the lines that are actually on the graph. That is, the level sets are points in the domain of the function above which the function is the same height.

Pick the first level set in the picture: suppose that the height of the function for every point on the level set is $\alpha$. Notice that for every point above and to the right of this level set, the value of the function at this point is larger than $\alpha$. Hence the set of points above and to the right of this level set is the upper contour set of the function corresponding to the value $\alpha$.

This is a source of endless confusion for everybody: compare the two curves in Fig. 11. The two curves are identical except for the labels. The interpretation of the curves is entirely different.

1. On the left, we have the graph of a function of one variable; area NE of the line is the area above the graph; area SW of the line is the area below the graph;
2. On the right, we have the level set of a function of two variables; area NE of the line is an upper contour set of the function; area SW of the line is an lower contour set of the function. In this case,
Where are the upper contour sets located in the left panel of the figure? Ans: pick $\alpha$ on the vertical axis. Find $x^\alpha$ on the horizontal axis that’s mapped to that point $\alpha$. The interval $[0, x^\alpha]$ is the upper contour set corresponding to $\alpha$.

Some familiar economic examples of level sets and contour sets.

1. level sets that you know by other names: indifference curves; isoprofit lines; budget line ($p \cdot x = y$).

2. lower contour sets that you know by other names: budget sets; production possibility set;

3. upper contour sets that you know by other names: think of the “region of comparative advantage” in an Edgeworth box: this is the intersection of the upper contour sets of the two traders’ utility functions.

Some practice examples for level sets.

- What are the level sets of a single variable function with no flat spots? Ans: A discrete (i.e., separated) set of points.
- What are the level sets of a concave single variable function with no flat spots? How many points can be in a level set? Ans: At most two.

- Now consider a function $f$ of two variables that has a strict local maximum at $x^*$ (i.e., $f$ is strictly higher at $x^*$ than on a nbd). What can you say about the level set of the function through $x^*$? Ans: The point $x^*$ must be an isolated point of the level set. Not necessarily the unique point, but certainly isolated.